JUN 2 7 1987

46-A-7

COMPLETED

ED 1.310/2:273073

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED273073

University of Minnesota

RESEARCH REPORT #7

ASSESSMENT PRACTICES IN MODEL EARLY CHILDHOOD EDUCATION PROGRAMS

Camilla A. Lehr, James E. Ysseldyke, and Martha L. Thurlow

EARLY CHILDHOOD ASSESSMENT PROJECT

ORIGINAL

April, 1986

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James E Ysseldyke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

DOCUMENT RESUME

ED 273 073 EC 190 268

AUTHOR	Lehr, Camilla A.; And Others
TITLE	Assesment Practices in Model Early Childhood Education Programs. Research Report #7. Early Childhood Assessment Project.
INSTITUTION	Minnesota Univ., Minneapolis.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE	Apr 86
GRANT	G008400652
NOTE	33p.; For research reports 9 and 10, see EC 182 736-741.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Demonstration Programs; *Disabilities; National Surveys; Preschool Education; *Testing; Test Reliability; *Test Use; Test Validity
IDENTIFIERS	*Handicapped Childrens Early Education Program

ABSTRACT

Information about current assessment practices was obtained from 54 surveys completed by Handicapped Children's Early Education Program (HCEEP) demonstration projects across the United States. Information about factors influencing the selection and continued use of tests also was provided. Results indicated that 19 tests were used by five or more programs and only one device was used by over half of the responding programs. Although most tests were listed as being used for more than one purpose, some tests appeared to be used more exclusively than others for a particular purpose. The technical adequacy of tests, in terms of norms, validity and reliability, was reportedly an important factor influencing selection and continued use. However, analysis of the 19 most commonly used devices revealed that only three were technically adequate. Other methods of assessment also were examined. The HCEEP Assessment Device Survey and a reference list of tests used are appended. (Author/CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEST COPY AVAILABLE

Abstract

Information about current assessment practices was obtained from 54 surveys completed by Handicapped Children's Early Education Program demonstration projects across the United States. Information about factors influencing the selection and continued use of tests also was provided. Results indicated that 19 tests were used by five or more programs and only one device was used by over half of the responding programs. Although most tests were listed as being used for more than one purpose, some tests appeared to be used more exclusively than others for a particular purpose. The technical adequacy of tests, in terms of norms, validity and reliability, was reportedly an important factor influencing selection and continued use. However, analysis of the 19 most commonly used devices revealed that only three were technically adequate. Other methods of assessment also were examined. Implications for model practice are discussed.

The development of this report was supported in part by Grant No. G008400652 from Special Education Programs, U.S. Department of Education. Points of view or opinions stated in this report do not necessarily represent official position of Special Education Programs.

BEST COPY AVAILABLE

Assessment Practices in Model Early Childhood Education Programs

Camilla A. Lehr, James E. Ysseldyke, and Martha L. Thurlow

The movement toward early childhood education is generally recognized as emerging in the 1960s (Lichtenstein & Ireton, 1984; Osborn, 1975). With the passage of the Economic Opportunity Act in 1964, extensive funding was provided for educational programs for preschool children (Ysseldyke & Algozzine, 1984). In 1968, Public Law 90-538 (Handicapped Children's Early Education Assistance Act) was passed, thus providing further support for the education of preschool handicapped children. Numerous experimental projects aimed at providing enrichment experiences and educational opportunities emerged in an attempt to demonstrate the beneficial effects of early education. The interest in early childhood education and assessment in the 1960s consequently spurred the development of many new marketable tests. A review of contemporary preschool asssessment instruments indicated that well over 200 tests were constructed and published in accordance with the Headstart movement (Dykes, Strickland, & Munyer, 1979).

Unfortunately, many of the early childhood tests that were developed were of poor quality. In 1971, the Center for the Study of Evaluation and the Early Childhood Research Center of the UCLA Graduate School of Education published a comprehensive guide of over 120 preschool and kindergarten tests (Hoepfner, Stern, & Nummedal, 1971). Of the 120 tests yielding 630 subtests, only 7 subtests were rated as providing good measurement validity. Most normative evaluations were either poor or fair.

BEST COPY AVAILABLE

Scrutiny of a test's technical characteristics is imperative because of the decisions that result from test scores. Guidelines for test construction and use, including standards for norms, validity and reliability, have been developed and outlined in the Standards for Educational and Psychological Tests (American Psychological Association, 1985). Unfortunately, the manuals of many tests lack sufficient information to justify their use for making decisions regarding young children. When Thurlow and Ysseldyke (1979) evaluated the validity, reliability and norms of the most frequently used tests in federally funded Child Service Demonstration Centers (CSDC), only 7 of the 28 tests were considered technically adequate in all three aspects.

It is evident that technical inadequacy of standardized tests for young children is a critical issue in assessment and educational decision making today. One of the major purposes of research investigating technical adequacy is to provide test consumers with information that enables them to choose and use a test in a more judicious and appropriate manner (Mardell-Czudnowski & Lessen, 1982). Evidence about the technical adequacy of tests can help to prevent selection of inappropriate or worthless measures.

Handicapped Children's Early Education Program

This study examined current assessment practices used in model programs on a national level. The Handicapped Children's Early Education Program's (HCEEP) demonstration projects were selected as the best source from which to gather this information because of the

BEST COPY AVAILABLE

national recognition they have received as model programs and the contributions they have made to the field of early education.

HCEEP was established in 1968 in an effort to provide major services and intervention to handicapped children at an early age (TADS, 1984). Its purpose is to "support experimental preschool and early childhood programs that show promise of promoting a comprehensive approach" to the problems of handicapped children and their families. The program was initiated by Congress with the passage of Public Law 90-538, the Handicapped Children's Early Education Assistance Act, and is supported by grants and contracts from the Office of Special Education Programs (OSEP) of the U.S. Department of Education.

A major part of the HCEEP (and the subject of this study) is its demonstration component. In 1983-84, the demonstration component consisted of 82 projects that developed and implemented innovative models of early intervention and education of young handicapped children. The model demonstration programs are composed of several features, including child identification and assessment, educational/therapeutic programming for children, evaluation of child progress, active parent and family participation, inservice training, coordination with public schools and other agencies, evaluation of project objectives, and demonstration and dissemination of project information (TADS, 1984). A recent report on an analysis of the impact of the demonstration and outreach components described their accomplishments as being "greater and more varied than for any other

BEST COPY AVAILABLE

documented education program identified" (Roy Littlejohn Associates, 1982).

Purpose

The purpose of this study was to determine what early childhood assessment instruments and other methods of assessment are being used in national model programs serving young handicapped children. Factors that influence the selection of tests and continued use of tests were also investigated. After gathering information about the most commonly listed assessment devices, the technical adequacy of each test was analyzed according to the purpose for which it was used.

Method

Subjects

The subjects of this study were 54 HCEEP demonstration projects located across the United States. The subjects were obtained from a pool of 82 demonstration projects in existence in 1983-84, as listed in the 1983-84 Handicapped Children's Early Education Program Directory. The Directory provides descriptive information about the HCEEP and was produced for the Office of Special Education Programs, U.S. Department of Education, by TADS in 1984. The 54 subjects represent those projects returning the survey sent to all demonstration projects; the number reflects a response rate of 65.9%.

Materials

A survey was developed to investigate issues related to current assessment practices used in the demonstration projects.

BEST COPY AVAILABLE

Specifically, personnel were asked to provide information about demographic characteristics of the project and factors influencing the selection and continued use of tests. The last section of the survey requested a list of actual tests, as well as other informal methods of assessment, used for five assessment purposes. The five purposes for administering tests, as developed by Salvia and Ysseldyke (1985), are defined as follows: (a) Screening: to identify students who are sufficiently different from others similar in age that they require special attention or assessment, (b) Classification/Placement: to identify students who are eligible for special education services, (c) Instructional Planning: to assist staff in planning educational programs (deciding what to teach and how to teach) for individuals, (d) Pupil Evaluation: to monitor individual progress, and (e) Program Evaluation: to evaluate the effectiveness of the educational program. A copy of the survey is included in Appendix A.

Procedure

The surveys were sent by mail to all 82 HCEEP demonstration projects funded during 1983-84, with cover letters and stamped envelopes enclosed for their return. A follow-up postcard was mailed approximately six weeks later to those centers not responding to the initial mailing.

Results

Project Information

Fifty-one demonstration projects provided information about the number of children served during 1983-84. Three of these projects

BEST COPY AVAILABLE

reported serving over 200 children each during 1983-84. These data were significantly discrepant from information reported in the 1983-84 edition of the HCEEP directory and were not used to calculate the number of children served. Compilation of data from the remaining 48 surveys indicated that the total number of children served during 1983-84 was 1,621. Within this sample the number of children served varied considerably among programs, ranging from 8 to 120 ($\bar{X}$ = 33.8, SD = 24.3).

For 53 projects, the average number of years of funding was 3.62 and ranged from 2 to 16 years. The age of children served ranged from prenatal care to 6 years. Thirty-eight programs (70.4%) provided services to children beginning at or before birth and 17 programs (31.5%) served children until the age of 6. The average ages of children served ranged from .80 for the youngest to 4.53. Age ranges appropriately reflected the early childhood population that the demonstration projects aim to serve.

Information provided about characteristics of the target population served was very general. The majority of respondents (53.7%) described the target population as having "various mild to moderate handicaps" without further specification. Twenty-seven percent described the children they served as being "at-risk." Only 14.8 percent described specific handicapping conditions of the target population served. These exclusive definitions included children who were specifically referred to as (a) language impaired (5.5%), (b) hearing impaired (3.7%), (c) behaviorally disordered (1.9%), (d)

BEST COPY AVAILABLE

severely/profoundly retarded (1.9%), and (e) visually impaired (1.9%). Because of the small number of children with specific handicapping conditions, tests were not analyzed according to the population served.

Factors Influencing Selection and Continued Use of Tests

Respondents were asked to select two factors (from a list of 10) that influenced the selection of tests used in the demonstration program. Table 1 is a summary of the factors influencing the selection of tests that were checked by respondents from the demonstration projects. Responses indicated that the most common factor influencing the selection of tests was whether the test was "recommended by other professionals" (64.8%). This was followed by "technical considerations (norms, reliability, and validity)" (61.1%). Twenty-two percent of the respondents indicated that availability or access to the test was an important factor. Approximately 18% selected inservice training workshops as influencing their selection of tests. Responses suggested that use of Tests in Print or Buros' Mental Measurements Yearbook, textbooks or journal articles, as well as publishers' catalogs and the cost of the tests, were relatively unimportant factors influencing test selection. Twenty-nine percent of the respondents checked the category "other" as influencing the selection of tests. Examination of the listed responses indicated several other factors that influenced test selection, including: (a) professional experience and expertise, (b) use of Educational Resources Information Center (ERIC), (c) graduate training with a

BEST COPY AVAILABLE

Table 1

Factors Influencing the Selection of
Tests Used in HCEEP Demonstraton Projects

Factor	Percentage of Programs Checking Factor
Recommended by Other Professionals	54.8
Technical Considerations (Norms, Reliability, Validity)	61.1
Availability or Access to Test	22.2
Inservice Training Workshops	18.5
Publisher's Catalog	5.5
Textbook	3.7
Use of Tests in Print or Mental Measurements Yearbook (Buros)	1.8
Cost	1.8
Journal	0.0
Other	29.6

BEST COPY AVAILABLE

particular instrument, (d) mandated use by county or state, (d) familiarity, and (f) compatibility with program objectives.

After a test has been selected, it is either discarded or continues to be used for various reasons. Respondents were asked to rate nine factors, in the order of their importance, influencing the continued use of tests. Forty-one respondents rated the factors as directed. Ratings were added together for each factor and averaged. "Information gathered from test results" received the highest average rating. The second most important factor was an "appropriate norming population." This was closely followed by the test's validity (third), reliability (fourth), ease of administration (fifth), and professional recommendations (sixth). Of lesser importance were the following factors: (a) common use of the test in the past (seventh), (b) favorable description by the test market (eighth) and (c) cost.

Assessment Devices

Fifty-two HCEEP demonstration projects provided information about the specific tests used to assess children and the purposes for which each test was used: The number of devices listed by each program varied, ranging from 1 to 16 ($\bar{X}$ = 6.2, SD = 3.8). A total of 109 tests was listed by 52 programs. Nineteen tests were used by five or more programs. (Seven projects listed unpublished project-developed tests, but each test was different.) No single test was used by every program. Only one test was used by over half of the responding programs -- Bayley Scales of Infant Development (52.8%). The specific devices used by five or more programs and the purposes for which they were used are summarized in Table 2.

BEST COPY AVAILABLE

Table 2

Percentages of HCEEP Model Demonstration Projects Using Assessment Devices for Different Purposes

Device[a]	% of Programs Using Device[b]	Purpose for Which Used[c]				
		Screening	Placement	Instructional Programming	Pupil Evaluation	Program Evaluation
Alpern-Boll	15.1	12.5	50.0	37.5	50.0	62.5
Bayley	52.8	10.7	71.4	14.3	25.0	42.9
Brigance	20.8	0.0	36.4	90.9	63.6	18.2
Denver	30.2	93.8	6.3	0.0	0.0	0.0
E-LAP	13.2	14.3	28.6	71.4	71.4	28.6
HELP	13.2	0.0	14.3	100.0	57.1	57.1
K-ABC	11.3	0.0	100.0	33.3	50.0	50.0
LAP	15.1	0.0	50.0	75.0	50.0	50.0
Leiter	9.4	0.0	80.0	0.0	20.0	20.0
McCarthy	18.9	10.0	90.0	20.0	20.0	30.0
Portage Guide	9.4	0.0	0.0	100.0	40.0	20.0
PPVT-R	13.2	14.3	85.7	28.6	28.6	28.6
REEL	9.4	20.0	60.0	0.0	40.0	40.0
SICD	20.8	0.0	72.7	45.4	45.4	27.3
Stanford-Binet	18.9	10.0	70.0	0.0	30.0	30.0
Uzgiris-Hunt	13.2	0.0	0.0	100.0	71.4	28.6
UPAS	9.4	0.0	20.0	100.0	60.0	80.0
Vineland	17.0	0.0	66.7	0.0	44.4	11.1
Zimmerman PLS	13.2	28.6	71.4	57.1	42.9	14.3
Project-Developed	13.2	57.1	0.0	42.9	28.6	28.6

[a]Full names of tests are provided in Appendix B.
[b]Percentages reflect numbers of HCEEP projects mentioning assessment device.
[c]Percentages reflect numbers of HCEEP projects using the device for each purpose based only on those listing the method.

BEST COPY AVAILABLE

It is apparent from Table 2 that programs using a particular test may have listed it for several different purposes. A summary of tests that were used for a particular purpose by 80% or more of the programs using it is provided in Table 3. Ninety-three percent of the programs that listed the Denver Developmental Screening Test listed it for the purpose of making screening decisions. No single test was used by 80% or more of the programs for the purpose of pupil evaluation. Of the programs listing the use of the Uniform Performance Assessment System (N = 5), 100% listed it for the purpose of instructional programming, and 80% (N = 4) listed it for the purpose of program evaluation. Other tests that were listed for instructional programming by at least 80% of the programs using the tests were: (a) Uzgiris-Hunt Ordinal Scales of Development (100%), (b) Hawaii Early Learning Profile (100%), and (c) Brigance Inventory of Early Development (90%). Tests that were used for making placement decisions (listed by at least 80% of the programs based only the number of programs listing the particular test) were: (a) Kaufman Assessment Battery for Children (100%), (b) McCarthy Scales of Children's Abilities (90%), and (c) Peabody Picture Vocabulary Test-Revised (86%).

The data on the purposes for which each assessment device is used reveal that nearly all devices are used for multiple purposes. All of the 19 tests listed by five or more programs were used for at least three purposes. However, many tests are used for some purposes more than for others. Tests that were listed five or more times within a particular category of purpose are listed in Table 4. The percentages

BEST COPY AVAILABLE

Table 3

Percentage of Programs Listing Tests for Each Purpose Based Only on the Number of Programs Citing the Device[a]

Screening	Classification/ Placement	Instructional Planning	Pupil Evaluation	Program Evaluation
Denver 93.8	K-ABC 100.0	HELP 100.0	--	UPAS 80.0
	McCarthy 90.0	Portage guide 100.0		
	PPVT-R 85.7	UPAS 100.0		
		Uzgiris-Hunt 100.0		
		Brigance 90.9		

[a]Table includes only those devices that were cited by 80% or more of the programs for a particular purpose based only on the programs citing the device. The device had to be cited by at least five programs to be included.

Table 4

Percentage of Programs Listing Tests for Each Purpose Based on the Total Number of Programs[a]

Screening	Classification/ Placement	Instructional Planning	Pupil Evaluation	Program Evaluation
Denver 28.3	Bayley 37.7	Brigance 18.9	Bayley 13.2	Bayley 22.6
	McCarthy 17.0	Uzgiris-Hunt 13.2	Brigance 13.2	
	SICD 15.1	HELP 13.2	E-LAP 9.4	
	Stanford-Binet 13.2	LAP 9.4	Uzgiris-Hunt 9.4	
	K-ABC 11.3	Portage guide 9.4	SICD 9.4	
	PPVT-R 11.3	SICD 9.4		
	Vineland 11.3	UPAS 9.4		
	Zimmerman PIS 9.4			

[a]Table includes only those devices mentioned by five or more programs. Percentages are based on the total number of programs.

BEST COPY AVAILABLE

listed in this table reflect the number of programs using a particular device out of all programs listing tests.

For screening, the most commonly used test, listed by 28.3% of the programs, was the Denver Developmental Screening Test. For the purpose of making placement decisions, eight tests were listed by five or more programs, with the most commonly used test being the Bayley Scales of Infant Development (37.7%). For the purpose of instructional programming, seven tests were listed by five or more programs. The most commonly listed test for instructional programming was the Brigance Inventory of Early Development (18.9%). For pupil evaluation, five tests were listed by five or more programs, with the most commonly used tests being the Bayley (13.2%) and the Brigance (13.2%). Last, for program evaluation, the most commonly listed test was the Bayley Scales of Infant Development (22.6%).

Respondents also were asked to list, according to the five categories of purposes, other methods that are currently used in their programs. Overall, the other methods of assessment that were used fell into 10 categories and included information gathered from: (a) parental involvement using interviews, questionnaires or consumer satisfaction measures, (b) observation, (c) teacher or staff input from questionnaires, meetings, or interviews, (d) individualized educational programs (IEP) development and reviews, (e) referrals or records including medical histories, (f) continuous monitoring and data collection, (g) videotapes of child's interaction, (h) home visits, (i) Family Needs Assessment, and (j) evaluations conducted by

BEST COPY AVAILABLE

an outside party. Nearly all of the respondents listed parental involvement (94.3%) and observational methods (83.0%) as other means of assessment. With the exception of outside evaluations, which were exclusively used for the purpose of program evaluation, all of the methods were listed as being used for at least three purposes. A summary of other methods listed as being used by HCEEP demonstration projects across purposes is presented in Table 5.

The average number of tests and the average number of other methods listed for each purpose are summarized in Table 6. For each purpose, it appears that both tests and other methods are used to help make decisions. However, results suggest that tests are more frequently used than other methods when making decisions about classification and placement, instructional planning, and pupil evaluation. This is most evident in the category of classification and placement ($t = 5.1$, $p < .01$).

Technical Considerations

Data on the commercial tests used for assessment by five or more projects (those listed in Table 2) were judged in terms of their technical adequacy. In order to be judged technically adequate, the test's norms, validity and reliability must all meet specified criteria. The criteria used in this study were compiled from several sources including the Standards for Educational and Psychological Tests (American Psychological Association, 1985), Assessment in Special and Remedial Education (Salvia & Ysseldyke, 1985), and an article by Mardell-Czudnowski and Lessen (1982). The criteria used

BEST COPY AVAILABLE

Table 5

Percentages of HCEEP Model Demonstration Projects Using Other Methods of Assessment for Different Purposes

Method	Percent of Programs Using Method[a]	Purpose[b]				
		Screening	Placement	Instructional Programming	Pupil Evaluation	Program Evaluation
Parental Involvement	94.3	56.0	14.0	24.0	14.0	56.0
Observation	83.0	59.1	38.6	38.6	36.3	20.5
Teacher/Staff Input	37.7	25.0	15.0	40.0	10.0	45.0
IEP Review	30.2	0.0	0.0	18.8	75.0	56.3
Referral Information/Records	20.8	72.7	27.3	18.2	0.0	0.0
Graphing/Data Collection	20.8	0.0	0.0	36.4	72.7	27.3
Videotapes	13.2	0.0	28.6	57.1	28.6	57.1
Home Visits	7.5	25.0	25.0	0.0	50.0	50.0
Family Needs Assessment	5.7	33.3	0.0	66.7	33.3	33.3
Outside Evaluations	5.7	0.0	0.0	0.0	0.0	100.0

[a]Percentages reflect numbers of HCEEP projects mentioning method of assessment

[b]Percentages reflect numbers of HCEEP projects using the method for each purpose based only on those listing the method

BEST COPY AVAILABLE

Table 6

Average Number of Tests and Other Methods Listed for Each Purpose

Purpose	Tests ($\bar{X}$)	Other Methods ($\bar{Y}$)	t
Screening	1.46	1.68	-1.0
Classification/Placement	2.48	.78	5.1*
Instructional Planning	2.83	1.31	3.9*
Pupil Evaluation	2.57	1.20	4.5*
Program Evaluation	1.87	1.46	1.4

*Significant at .01 or less

BEST COPY AVAILABLE

for evaluating each test's norms, reliability and validity are presented in Table 7. All of the tests except the Portage Guide to Early Education and the Uzgiris Hunt Ordinal Scales of Development were analyzed in light of their use as an instrument to help make classification and placement decisions, which requires the most stringent reliability coefficients. Tests that were specifically described as criterion referenced (although they may have given some age guidelines) were not analyzed with respect to the technical adequacy of their norms because they presumably are not used to make normative comparisons. Only information contained in the most current test's manual was used to analyze each instrument's technical adequacy.

The tests used by the model demonstration projects are evaluated in Table 8. The evaluation indicated that of the 19 instruments used by five or more programs, only three were technically adequate on all three dimensions (using most stringent reliability criteria as dictated by the purpose for which the test is used by the model program).

Discussion

This study investigated current assessment practices of HCEEP demonstration projects. Many studies have documented the educational contributions and effectiveness of HCEEP demonstration projects, but none have comprehensively examined the assessment practices actually used in these model programs.

The selection of tests that are used in HCEEP demonstration models reportedly is based largely on recommendations by other

BEST COPY AVAILABLE

Table 7

Criteria for Evaluating Technical Adequacy of Tests

Norms	Reliability	Validity
1. Norms should be available in manual or in an accompanying technical publication.	1. The manual should present evidence of reliability. Although the manual should contain the reports on reliability, additional sources such as technical reports, should be consulted.	1. Validity should be reported in the manual or in an accompanying technical publication.
2. Norms should be clearly defined and describe the populations especially for comparative purposes. Such defining characteristics of populations should include: the age(s), grade level, gender, geographic regions used, race, and handicapping conditions found within the norming population.	2. Reliability coefficients as well as standard errors of measurement should be presented in a tabular format.	2. Evidence of at least one type of validity should be presented for the major types of inferences for which the use of a test is recommended (i.e., criterion-related; concurrent or predictive; content; construct).
3. The norm-sampling method should be well defined.	*3. Reliability procedures and samples should be described.	3. For content validity, the manual should define the content area(s). Tests that are based on content validity should update content in revised forms.
4. The norm-sampling should not have been based upon convenience or readily available populations.	4. At least one type of reliability used should be stated (i.e., test-retest, alternate form, internal inconsistency, split-half, interrater reliability).	4. For construct validity, the manual should clearly define the ability or aptitude measured. For tests for which there is a time limit, the manual should state how speed affects scores.
5. Revised tests should provide norm comparisons for all forms. It should be noted that criterion-referenced tests do not need to present norming information (Popham, 1971).	5. For making decisions regarding individuals, reliability coefficients must be greater than or equal to .90 (e.g., instructional planning and placement.	5. For both types of criterion-related validity (a) the criteria should be defined; (b) validity should be reported; (c) samples should be completely described; (d) correlation coefficients with other tests should be reported; and (e) for predictive validity, a statement concerning the length of time for which predictions can be made should be included.
6. One hundred should be the minimum number of persons in any norm sample per age or grade (Salvia & Ysseldyke, 1985).	6. For screening decisions, reliability coefficients of .80 are acceptable.	
7. In assessing individuals with handicapping conditions, the test user should use either regular or special norms, depending on the purpose of the testing.	7. For administrative purposes and group decisions, reliability of .60 is acceptable.	
8. The test's norms should not be older than 15 years (Ysseldyke & Algozzine, 1984).		

*In this analysis, reliability studies with sample sizes less than 25 were not considered.

BEST COPY AVAILABLE

Table 8

Technical Adequacy of Devices Used by Five or More HCEEP Demonstration Projects

Device	Norms	Reliability			Validity
		≥.90 Individual	≥.80 Screening	≥.60 Group	
Alpern-Boll (1972)	+	-	-	-	+
Bayley (1969)	-	-	+	+	-
Brigance (1978)	CR	-	-	-	+[a]
Denver (1975)	-	-	-	-	+
Early LAP (1975)	CR	-	-	-	+[a]
HELP (1979)	CR	-	-	-	+[a]
Kaufman (1983)	+	+	+	+	+
LAP (1981)	CR	-	-	-	+[a]
Leiter (1979)	-	-	-	-	-
McCarthy (1972)	+	+	+	+	+
Portage (1976)	CR	-	-	-	+[a]
PPVT-R (1981)	+	-	+	+	+[a]
REEL (1971)	-	-	-	-	-
SICD (1975)	-	-	-	-	+[a]
Stanford-Binet (1972)	-	-	-	-	-
UPAS (1981)	*	*	*	*	*
Uzgiris-Hunt (1975)	-	-	-	+	+[a]
Vineland (1984)	+	+	+	+	+
Zimmerman PLS (1969)	-	-	-	-	-

[a]Evidence of content validity only is based on information contained in test manuals only.
Ratings in table are: + technically adequate, - technically inadequate, * manual not available.

BEST COPY AVAILABLE

professionals and technical considerations. Written material from textbooks, journals or publisher catalogs does not appear to be of utmost importance when tests are selected. However, inservice training workshops appear to have some impact on whether a test is selected for use. This information has implications for those who develop and market assessment instruments, suggesting that the chances of a test being selected are increased if it is technically adequate and has been used and recommended by other professionals in the field.

The continued use of tests in HCEEP model demonstration projects is most strongly influenced by the information gathered from the test's results. This makes intuitive sense and one would hope that HCEEP model programs are using tests that provided useful information. The next most important factors influencing the continued use of tests were the tests' norms, validity and reliability. Again, it appears that HCEEP projects are adhering to guidelines which document the importance of using technically adequate devices (American Psychological Association, 1985; Salvia & Ysseldyke, 1985; Ysseldyke & Algozzine, 1984). However, examination of the tests that are being used reveals that, in many cases, HCEEP Model Demonstration projects are using devices that are technically inadequate.

Over 100 tests were listed by the 54 projects surveyed. Nineteen tests were used by five or more programs and only one test was used by over half of the responding programs. All of the tests listed were analyzed according to the purpose for which they were used. Nearly all of the 19 tests listed were used for several purposes. However,

BEST COPY AVAILABLE

some tests were used more exclusively for some purposes than for others. For example, the Denver Developmental Screening Test was generally used for screening. For making placement and classification decisions, the Kaufman Assessment Battery for Children, McCarthy Scales of Children's Abilities and Peabody Picture Vocabulary Test-Revised were primarily used. Five tests (HELP, Portage Guide, UPAS, Uzgiris-Hunt, and Brigance) were used particularly for instructional planning and one was particularly cited for program evaluation (UPAS).

Results suggested that no single test was used for pupil evaluation more exclusively than others. Perhaps the demonstration programs relied more heavily on informal data-based measurement systems to monitor individual student progress. Such measures usually are tied to the curriculum, simple to administer, reliable, valid and sensitive to small fluctuations in student performance (Ysseldyke, Thurlow, Graden, Wesson, Algozzine, & Deno, 1983). Yet, some model programs continue to use tests that yield IQs for measuring pupil progress -- tests that are clearly inappropriate (Salvia & Ysseldyke, 1985). Examination of the data indicated that nearly 21% of the programs listed continuous graphing and data collection as another method of assessment. Of those who listed this measurement technique, 72% used it for pupil evaluation. However, tests were still listed more frequently than other informal methods as a means of evaluating or monitoring student progress.

Of the 19 tests most frequently cited, and analyzed according to the purpose for which they were used (using the most stringent

BEST COPY AVAILABLE

reliability criteria, r $\geq$ .90), only three had technically adequate norms, validity and reliability. Although results suggested that tests are selected and continue to be used based largely on their technical adequacy, many of the tests used are not technically adequate. Practitioners might explain this discrepancy by pointing out that these are the only tests available. Nevertheless, using a technically inadequate device cannot be justified or excused, because of the important decisions that are made based on the data gathered from such devices.

The criteria used for analyzing a test's technical adequacy in this study were determined by guidelines provided by the APA (1985); and research conducted by Ysseldyke and Algozzine (1984); and Mardell-Czudnowski and Lessen (1982). Although a test user can be more confident of a test's technical adquacy based on these criteria, it is essential for test users to examine qualitative aspects of norms, reliablity and validity in addition to quantitative indexes. For example, sample size in reliability studies must be considered, research studies investigating validity that do not appear in the manual perhaps should be considered, the test's accuracy in making correct decisions might be examined, and the test's purpose as designated by its authors must be examined. Ultimately, it is the test user's responsibility to determine the value of a test based on documented research.

Fortunately, it looks as though decisions that are made about children in HCEEP programs are based on more than one test. In

BEST COPY AVAILABLE

addition, other methods are used for the assessment of young children. Nearly all of the projects used information gathered from parents and observations in their assessment process. Other sources of information included teacher/staff input, IEP review, continuous graphing and data collection, and home visits. These techniques become especially important when the inadequacy of the tests used is considered. Although the use of these methods appears to be secondary and supplementary to the use of tests in making decisions, the HCEEP model demonstration programs may be shifting toward a comprehensive process of data gathering for making educational decisions.

HCEEP projects have developed many products to assist in the assessment of young children. As model early childhood programs, they are charged with the responsibility of using and developing sound assessment practices. Perhaps among the 100 tests that were listed by the 54 projects surveyed, some of the less frequently used project developed tests hold promise in providing useful information and being technically adequate. If so, it is important to disseminate this information to wider audiences who are in need of sound devices on which to base early childhood decisions. In addition, the awareness and use of other methods that are used to contribute to a comprehensive process of data gathering are of critical importance in making decisions about young handicapped children.

BEST COPY AVAILABLE

References

American Psychological Association. (1985). Standards for educational and psychological tests. Washington, DC: APA.

Dykes, J. K., Strickland, A. M., & Munyer, D. D. (1979). Assessment and evaluation instruments for early childhood programs. Gainesville, FL: Florida Educational Research and Development Council (ERIC Document No. ED 171378).

Hoepfner, R., Stern, C., & Nummedal, S. G. (1971). CSE-ERIC preschool/kindergarten test evaluation. Los Angeles: UCLA Graduate School of Education.

Lichtenstein, R., & Ireton, H. (1984). Preschool screening: Identifying young children with developmental and educational problems. Orlando, FL: Grune & Stratton, Inc.

Mardell-Czudnowski, C. D., & Lessen, E. I. (1982). Technical adequacy of assessment instruments: Can we agree? Diagnostique, 4, 189-202.

Osborn, D. K. (1975). Early childhood education in historical perspective. Athens, GA: Education Associates.

Popham, J. W. (1971). Criterion-referenced measurement: An introduction. Englewood Cliffs, NJ: Educational Technology.

Roy Littlejohn Associates, Inc. (1982). An analysis of the impact of the handicapped children's early education program final report. Washington, DC: Author.

Salvia, J., & Ysseldyke, J. E. (1985). Assessment in special and remedial education (3rd ed.). Boston: Houghton-Mifflin.

TADS (Technical Assistance Development System). (1984). Handicapped children's early education program. D. Assael (Ed.). Chapel Hill, NC: Office of Special Education Programs, U.S. Department of Education.

Thurlow, M. L., & Ysseldyke, J. E. (1979). Current assessment and decision-making practices in model LD programs. Learning Disability Quarterly, 2, 15-24.

Ysseldyke, J. E., & Algozzine, B. (1984). Critical issues in special and remedial education. Boston: Houghton-Mifflin.

Ysseldyke, J. E., Thurlow, M. L., Graden, J. L., Wesson, C., Algozzine, B., & Deno, S. L. (1983). Generalization from five years of research on assessment and decision making: The University of Minnesota Institute. Exceptional Education Quarterly, 4(1), 75-93.

BEST COPY AVAILABLE

Appendix A

HCEEP Assessment Device Survey

BEST COPY AVAILABLE

HCEEP ASSESSMENT DEVICE SURVEY

1. Center Information

a. Years of funding: 19___ to 19___

b. Age range of children served: _____ to _____

c. Number of children served in 1983-84: _____

d. Target population(s) served: _____

_____

2. Tests are selected in various ways. We are interested in knowing what factors influence the selection of tests you use in your program. Please check the two factors that most often apply.

____ Inservice training workshops
____ Recommended by other professionals
____ Use of *Tests in Print* or editions of *Mental Measurements Yearbook* (Buros)
____ Publisher's catalog or advertisements describing test
____ Availability or access to test
____ Technical considerations (norms, reliability, validity)
____ Textbook
(name) _____
____ Journal article
(name) _____
____ Cost of test
____ Other (please list) _____

3. Tests are used for various reasons. How important are the factors listed below in influencing the use of the most commonly used tests in your program? Please rate each in order of importance (1 = most important, 2 = next most important, . . . 9 = least important).

____ Easy to administer
____ Common use of test in the past
____ Appropriate norming population
____ Adequate reliability as specified by APA standards (1972)
____ Adequate validity as specified by APA standards (1972)
____ Information gathered from test results
____ Recommended by other professionals
____ Cost of test
____ Favorably described by test market

~~29~~ 26

BEST COPY AVAILABLE

- OVER -

4. Tests are administered for a variety of purposes. For each purpose defined below, list the tests and any other methods that your program currently uses.

PURPOSE	TESTS USED *(e.g., McCarthy Scales of Children's Abilities)*	OTHER METHODS *(e.g., parent interview, observation)*
SCREENING To identify students who are sufficiently different from others similar in age that they require special attention or assessment.		
CLASSIFICATION/ PLACEMENT To identify students who are eligible for special education services.		
INSTRUCTIONAL PLANNING To assist staff in planning educational programs (deciding what to teach and how to teach) for individuals.		
PUPIL EVALUATION To monitor individual progress.		
PROGRAM EVALUATION To evaluate the effectiveness of the educational program		

BEST COPY AVAILABLE

Appendix B

Reference List of Tests Used

BEST COPY AVAILABLE

Reference List of Tests Used

Alpern-Boll	Alpern, G., & Boll, T. J. (1972). Developmental profile. Aspen, CO: Psychological Development Publication.
Bayley	Bayley, N. (1969). Bayley scales of infant development. New York: Psychological Corporation.
Brigance	Brigance, A. (1978). Brigance inventory of early development. USA: Curriculum Associates, Inc.
Denver (DDST)	Frankenberg, W. K., Goldstein, A., & Camp, B. W. (1975). Denver developmental screening test: Revised reference manual. Denver, CO: LADOCA Foundation.
E-LAP	Glover, M., Preminger, J. L., & Sanford, A. (1975). Learning accomplishment profile for infants (Early LAP). Winston Salem, NC: Kaplan Press.
HELP	Enrichment Project for Handicapped Infants. (1979). Hawaii early learning profile. Palo Alto, CA: VORT Corporation.
K-ABC	Kaufman, A. S. & Kaufman, L. (1983). Kaufman assessment battery for children. Circle Pines, MN: American Guidance Service, Inc.
LAP	Sanford, A. R. (1974). (1981). A manual for use of the learning accomplishment profile. Winston-Salem, NC: Kaplan School Supply.
Leiter	Leiter, R. G. (1979). The Leiter international performance scale. Chicago, IL: Stelting Co.
McCarthy	McCarthy, D. (1972). Manual for the McCarthy scales of children's abilities. New York: Psychological Corporation.
Portage	Bluma, S., Shearer, M., Frohman, A., & Hillard, J. (1976). Portage guide to early education. Portage, WI: Portage Project Cooperative Educational Service Agency.
PPVT-R	Dunn, L. M. & Dunn, L. (1981). Peabody picture vocabulary test - revised. Circle Pines, MN: American Guidance Service.

BEST COPY AVAILABLE

REEL — Bzoch, K., & League, R. (1971). REEL manual, assessing language skills in infancy. Baltimore, MD: University of Park Press.

SICD — Hedrick, D., Prather, E., & Tobin, A. (1975). Sequenced inventory of communication development. Seattle: Univeristy of Washington Press.

Stanford-Binet — Terman, L. M. & Merrill, M. A. (1973). Stanford Binet intelligence scale. Boston, MA: Houghton-Mifflin.

UPAS — Haring, N. G., White, O. R., Edgar, E. B., Afflick, J. Q., & Hayden, A. H. (1981). Uniform performance assessment system - Examiner's manual. Columbus, OH: Charles E. Merrill.

Uzgiris-Hunt — Uzgiris, I. C. & Hunt, J. MCV. (1975). Assessment in infancy: Ordinal scales of psychological development. Urbana: University of Illinois Press.

Vineland — Sparrow, S., Balla, D., & Cicchetti, D. (1984). Vineland adaptive behavior scales. Circle Pines, MN: American Guidance Service, Inc.

Zimmerman PLS — Zimmerman, I., Steiner, V., & Evatt, R. (1969). Preschool langauge scale. Columbus, OH: Charles E. Merrill Publishing Co.

BEST COPY AVAILABLE